CHEMICAL STRUCTURE ELUCIDATION OF AN UNKNOWN COMPOUND BY DIRECT ANALYSIS IN REAL-TIME MASS SPECTROMETRY AND NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Suckwon Lee

August 2019

ABSTRACT

Direct Analysis in Real-Time Mass Spectrometry (DART-MS) and Nuclear Magnetic Resonance (NMR) are two analytical techniques that are most used in structural chemistry and biology. DART-MS measures mass to charge ratio of different compounds in a solution. From the mass to charge ratio, the chemical formula of a substance can be deduced. NMR can determine chemical structure of a purified substance. Thus, with the chemical information of molecules gathered by DART-MS and NMR, chemical structures of the molecules can be unambiguously determined.

In this thesis, the chemical structure of an unknown compound was found by DART-MS and NMR. The unknown compound was dissolved in chloroform ($CDCl_3$) solvent, and exact mass of the unknown compound was measured by DART-MS. After the mass measurement, the solution was run in NMR for the ^1H NMR and 2D NMR. With the data from DART-MS and NMR structure, the chemical structure of the compound was determined to be nuciferine.

BIOGRAPHICAL SKETCH

Suckwon Lee was born and raised in Seoul, Korea. Suckwon attended De Anza college in Cupertino, California. He transferred to University of California, Davis where he got his degree of bachelor of science in molecular biology. When he was senior in Davis, he also interested in computational biology, so he joined the lab of Dr. Sato in Davis's Department of Pharmacology for 7 months and learned computational techniques in biology. After the graduation from Davis, he wanted to broaden his skill set so that he might study other aspects. He finally came to Cornell as a Masters of Chemistry student and has learned chemical techniques more deeply than before. Following his all studies, he plans to use all the techniques and knowledge for his own unique research.

ACKNOWLEDGMENTS

Thanks to my family for supporting and encouraging me.

Thanks to my friends for standing by my side.

Thanks to all the professor, who taught me, in Cornell University.

Thanks to David, Ivan, and Frank for everything.

I thank you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 MASS SPECTROSCOPY

1.1.1 Elucidation of chemical structure

MS is a powerful analytical technique that measures mass to charge ratio of positively or negatively ionized compounds in many different fields of chemistry. The intensity of the mass to charge ratio can be different from the components of compounds. This result is shown as different patterns of mass spectrum. The exact mass can deduce combinations of possible chemical formulae, and the different patterns of mass spectrum can narrow the possible chemical formulae by isotopic signature, which is a specific intensity ratio pattern of specific atom components in a compound. In addition, from the deduced chemical formulae, the possible ring double bond equivalents (RDBE) can be determined. The ionized compound can also be fragmented or adducted form from different chemical structures, making specific pattern of mass spectrum in lower or higher mass than its original mass. Thus, MS helps to elucidate a chemical structure of a target compound through the combination of information with RDBE from the most possible chemical formula and its fragmentations.

1.1.2 Direct analysis in real time (DART)

To measure the mass to charge ratio of compounds, the compounds must be positively or negatively ionized, and there are many types of ionization techniques, including Direct Analysis in Real Time (DART), a type of ambient ionization[1,2]. DART works at atmospheric pressure. It uses a heated gas flow of neon or nitrogen to produce excited neon or nitrogen gas at the main chamber. The energy of the excited gas is transferred

to the analyte to make ionized compounds. This ionization avoids the sample to directly expose high voltages, laser beams, or plasma. In its key characteristics, this ionization mechanism is similar to that of electrospray ionization (ESI).

The mixture of ions produced in MS must be separated into a collection of mass-specific ions by other techniques which combined with MS for high sensitivity, selectivity, and specificity. Despite the fact that gas chromatography (GC) and liquid chromatography (LC) are widely used for the separation, they require tedious sample preparation and long analysis time. DART can be also combined with the chromatography, but the combination removes the benefits of high rapidness and simplicity of DART. DART has been successfully used for analysis of compounds in any states (solids, liquids, or gases) without pre-treatment of sample preparation. In addition, DART has no requirement of using chromatography, allowing for rapid screening of samples. Because of the reasons, the technique is called direct analysis in real time.

1.1.3 Orbitrap

After ionization and separation of ions, a mass analyzer is required for mass selection. Ions should be separated by their mass to charge ratio for the specificity. There are many types of mass analyzers including the orbitrap, which is a type of ion traps[3,4]. Orbitrap uses an electrostatic field to trap ions in orbit around an electrode, and each ion produces an oscillating field at a specific frequency of each type of ions. The oscillating charges are detected and their frequencies are Fourier transformed to be recorded. This technique has high sensitivity and high resolving power.

1.2 1-D NMR

1.2.1 Magnetization and Larmor frequency

Nuclear magnetic resonance (NMR) is commonly used for solving more definitively the chemical structure of substances. NMR is a phenomenon where nuclei are perturbed by an external magnetic field (B_0) and electromagnetically signal back as at a characteristic frequency. Each nucleus has the quantum property of spin with vector characteristic. When a B_0 is applied to the nuclei, non-zero spin value of nuclei is aligned with the B_0, creating net equilibrium magnetization along with B_0. After perturbation of the alignment by a specific radio frequency (rf), the force between the B_0 and the tilted nuclear spin axis creates a precession of the nuclear spin around the B_0. The magnetic precession takes place at an isotope-specific frequency, called the Larmor frequency. Usual target nuclide in organic compound are ^{1}H, ^{13}C, and ^{15}N, all of which have a nuclear spin of one-half.

1.2.2 Pulse

To manipulate or maximize the observation of frequency of precession, an electromagnetic field called B_1 is applied in the direction perpendicular to B_0. B_1 can be turned on and turned off by using appropriate time and strength of radio frequency (RF). If B_0 is directed on Z-axis in XYZ coordinate system, B_1 is applied to either the Y- or X-axes to make the spin vector of nuclei rotate about the Y- or X-axes. The time duration the B_1 pulse required to rotate the spin vector exactly 90 degrees is said to be a 90-degree pulse. Most often the field strength is calibrated by observing the signal after a 180-degree pulse, which is more readily optimized as the observed signal is a minimum.

1.2.3 Relaxation time

Any magnetic moment excited by pulses ultimately relaxes back to equilibrium along the Z-axis, producing no signal on X- or Y-axes. There are two types of relaxation: spin-lattice and spin-spin. Spin-spin relaxation is called transverse or T_2 relaxation, which is the relaxation of the excited magnetic moment perpendicular to B_0. The T_2 relaxation affects the spectral line-width. Spin-lattice relaxation called longitudinal or T_1 relaxation is the return to equilibrium in the direction of B_0. Because the measurement of Larmor frequency is measured on X- or Y-axis, both of T_1 and T_2 relaxation must be considered for the cycle of frequency measurements.

1.2.4 FID

After using 90-degree pulse, the magnetization vector perpendicular to the B_0 precesses during its relaxation time, and the vector returns back to equilibrium. This process is called free induction decay (FID), and the FID signal depends on the t_2 relaxation because of the detector catches the transverse frequency of FID.

1.2.5 Shielding and deshielding

In addition to the B_0, the surrounding electrons of nuclei can also perturb the frequency of precession. Electrons near nuclei located in a magnetic field rotate giving an opposite direction of magnetic field against the B_0. Because of the magnetic field produced by electrons near nuclei, the effective magnetic field at a nucleus is reduced, producing less angular frequency. The effect of the electron distribution decreases the nuclear frequency, leading to a shift upfield (to lower chemical shift on the common δ scale) is called shielding. In contrast, when electron density is lowered, the nucleus feels stronger

magnetic field. This effect is called deshielding. For example, high electronegativity atoms such as F, O, and N attached to proton pull electrons from the proton, and the proton becomes deshielded. There can be also anisotropic magnetic field by circulating π-electrons in aromatics that produce induced magnetic field. Depending on the position of nuclei, nuclei feels more magnetic field or less magnetic field due to the secondary magnetic field of π-electrons in aromatics Therefore, these shielding and deshielding deduce components of chemical groups in a molecule by using ^{1}H-NMR.

1.2.6 Spin-spin coupling

Nuclei also experience other magnetic field by other adjacent chemically non-equivalent nuclei. Spin of nuclei can be either up direction or down direction. This different direction of spin at adjacent nuclei can perturb the angular frequency of precession of target nuclei. This perturbation is usually occurred through bond in range of 3 or 4, and it is said spin-spin, scalar, or J coupling. For example, protons in methyl group attached to methine feel either up-direction or down-direction of a spin of chemically non-equivalent proton in methine by near electrons through bonds. The up-direction spin aligned with B_0 deshields the protons in methyl otherwise the down-direction spin shields them. The result is two split lines (doublet) at chemical shift of methyl protons in NMR spectrum. This use of coupling helps to decide how many protons are near target protons.

1.3 2-D NMR

1.3.1 Basic principles of 2-D NMR

In 1-D NMR, only 90-degree pulse is required right before the acquisition time. In 2-D NMR, evolution and mixing are required between preparation and acquisition time.

During the evolution time (t_1), the frequency of first nuclei evolves freely, and the magnetization is transferred to another second nuclei in the mixing time. The transferred magnetization on second nuclei is measured in acquisition time (t_2). Thus, in 2-D NMR, the magnetization transfer is measured. This magnetization transfer can be manipulated by different pulse sequences and used for many purposes.

1.3.2 ^{1}H-^{1}H COSY

^{1}H-^{1}H Correlation Spectroscopy (COSY) uses the transfer of magnetization on the proton to the second proton, based on their J-coupling. Thus, COSY is to assign the coupled proton chemical shifts. This quick and reliable technique can detect very small couplings that can be hardly detected in ^{1}H-NMR. The cross-peaks represent the proton coupling correlation.

1.3.3 ^{1}H-^{13}C HSQC

^{1}H-^{13}C Heteronuclear Single Quantum Correlation (HSQC) indirectly detects the protonated carbon chemical shifts by the transfer of magnetization on the proton to the attached carbon, based on their J-coupling. Thus, the purpose of using ^{1}H-^{13}C HSQC is to determine the chemical shift of ^{13}C sites which have bound 1H atoms. Furthermore, the HSQC signal helps clarify the multiplicity of attached protons attached to specific carbons, because the signal phase is inverted in methylene groups (CH_2) with an even number of attached H atoms, and methine (CH) or methyl (CH_3) groups with odd numbers of protons attached to carbons. The HSQC has largely replaced the older HMQC experiment as, this pulse sequence minimizes ^{13}C magnetization loss due to relaxation, and gives clearer peaks in F1 by removing ^{1}H-^{1}H coupling. However, careful

attention to pulse duration is required because this technique is sensitive to imperfect pulses.

1.3.4 ^{1}H-^{13}C HMBC

^{1}H-^{13}C Heteronuclear Multiple Bond Correlation (HMBC) is similar to ^{1}H-^{13}C HSQC in that it correlates proton and ^{13}C shifts, but the HMBC is optimized to assign longer-range couplings associated with atoms separated by 2-3 bonds. The 3-bonds correlations are usually stronger in intensity, while the 1-bond correlation can be detected but with low intensity. This is very useful to determine the structure of a molecule because the long-range correlation can clarify the connectivity. The lowered sensitivity of the HMBC can miss some carbons, and the signal of carbons can be not resolved with small chemical shift separation.

1.3.5 ^{1}H-^{1}H ROESY

Nuclear magnetic moment is also affected by other nuclei through space via the dipole-dipole interaction. This interaction is different from the scalar coupling. When a proton is magnetically saturated, another near proton is affected by the saturated proton, and relaxation occurs between the two protons. This relaxation changes the integrated intensity (positive or negative) of the near proton. The effect of intensity change is called Nuclear Overhauser Effect (NOE). However, in mid-sized molecules, the NOE is small due to the cancellation of positive and negative NOE. Thus, another technique, Rotating frame NOE (ROE)[5], is required to measure NOE in mid-sized molecules. Magnetization along transverse plane is spin locked by an external magnetic field, and the ROE is always positive.

Rotating Frame Nuclear Overhauser Effect Spectroscopy (ROESY) uses ROE. Thus, the purpose of using ROESY is to determine stereochemistry of a mid-sized molecule. The phase system of ROESY must be considered in careful manner. ROE cross-peaks are opposite phase to the diagonal. Same phase of cross-peaks with the diagonal represents chemical exchange, and mixed phase cross-peaks are TOCSY type artifacts which shows correlations between all protons, not just between geminal or vicinal protons as in COSY.

1.4 DATA PROCESSING

NMR signal strength depends on the transition of spins from lower to higher energy state. Because of the energy gap is so low in quantum spin system of nuclei, their population difference between lower and higher energy state is very similar. Although increasing the external magnetic force or NOE can increase the energy gap for making the population difference higher, there is also limitation due to the rapid increase of price for higher field magnets. Thus, computing processes can be required for improving the sensitivity. One common way to increase the sensitivity is to increase the signal to noise (S/N) ratio by manipulating the FID. At the tail of FID, it has signal, but also noise information. By applying negative exponential equation to the FID, the tail of FID can reach zero. However, because the short t_2 relaxation decrease the resolution, the linewidths of peaks can be broadened.

The FID can also be manipulated for better resolution. The idea is just same as the sensitivity modification. Thus, using combination of weighting equations, the FID can be modified on purpose: better sensitivity or resolution.

FID is represented by a series of data points, and the spectrum that Fourier transformed is a pack of data points. In other words, more data points can make more smooth lines. Thus, before modifying the FID, the data points of the original FID is usually doubled or even more by adding zeroes. This processing is called zero filling.

CHAPTER 2

EXPERIMENTAL

2.1 PREPARATION OF SAMPLE

The ~5mg of unknown sample in solid state was dissolved in ~5 ml of chloroform.

2.2 DART-ORBITRAP MS

To deduce the chemical formula of the unknown compound, DART-orbitrap was used. After measured background spectrum with a glass rod, a tip of glass rod was dipped into the solution, and mass to charge ratio of the solution in positive charge mode was measured in real time by Thermo Scientific Exactive DART-Orbitrap.

2.3 VARIAN 600 NMR

2.3.1 Preparation

The solution of unknown sample with chloroform was transferred into a Shigemi tube. Sample positioning was corrected by using a depth gauge, and the sample was inserted.

2.3.2 NMR experiments

^{1}H-NMR experiment was set up for assigning all protons chemical shifts. The sample was auto-shimmed, and 90-pulse was optimized in this experiment. In the second experiment, ^{1}H-^{1}H COSY experiment was set up for assigning the coupled protons, and the gradient COSY sequence was used. In the third experiment, ^{1}H-^{1}H ROESY experiment was set up for determining the stereochemistry, and the ROESYAD sequence was used. In the fourth experiment, ^{1}H-^{13}C HMBC experiment was set up for assigning the protonated carbon chemical shift, and the HSQCAD sequence was used. For the final experiment, ^{1}H-^{13}C HMBC experiment was set up for determining the connectivity of protons, and the gradient HMBCAD sequence was used.

2.4 XCALIBUR

The saved MS file was opened by Xcalibur software. Mass tolerance 5.0 ppm was used for the elution spectrum. In the display option of mass spectrum, 5 decimals were used.

Several peaks from the elution spectrum were determined if the peaks came from other chemicals other than the unknown sample by using range option. Because of the components of unknown sample were not known, formulae from masses of noticeable peaks from mass spectrum were generated by using all atom with 5 ppm error.

2.5 NMR DATA PROCESSING (MNOVA)

2.5.1 H-NMR

Zero filling was to 64k points, auto-baseline correction was used, and window function was applied for determining multiplicities. Three proton peaks were not resolved well to determine their multiplicities: A protons at 8.35, 7.24, 2.67 ppm. For the proton peaks, exponential function was used: -0.2 for 8.35 ppm, -0.5 for 7.24 and 2.67 ppm (In the Mnova, the negative exponential is actually upward).

2.5.2 COSY

The COSY experiment was failed. There were only three diagonal peaks observed. The reason was not known.

2.5.3 HSQC

Zero filling was to extended the data set to 4k points in F2 and 2k points in F1. Window functions were a Gaussian (7.65 GB[Hz]) in F2 and a sine square (90 degree) in F1. The bitmap plotting method with a and red-blue (gradient) palette was used for purposes of display.

2.5.4 HMBC

Zero filling was to 4k points in F2 and extended the data set to 8k points in F1. Window functions were Sinebell II (0.0%-50%) and Gaussian (3.82 GB[Hz]) in F2 and a Sine square (90 degree) in F1. The bitmap plotting method with a custom 4 palette were used for purposes of display. Adjusted magnification was used for each peak to determine the correlation.

2.5.5 ROESY

Zero filling was to 4k points in F2 and 2k points in F1. Window functions were a Gaussian (3.82 GB[Hz]) in F2 and a Gaussian (10.00 GB[Hz]) in F1. The Bitmap plotting method and red-blue (gradient) palette were used for purposes of display. Adjusted magnification was used for each peak to determine the ROE.

CHAPTER 3

RESULTS AND ANALYSIS

3.1 H-NMR

3.1.1 Basic information from ^{1}H-NMR

^{1}H-NMR spectrum peaks and the single chloroform peak at 7.25 ppm show reasonable quality of shimming[6]. The peaks are auto-integrated, excluding the three peaks at around

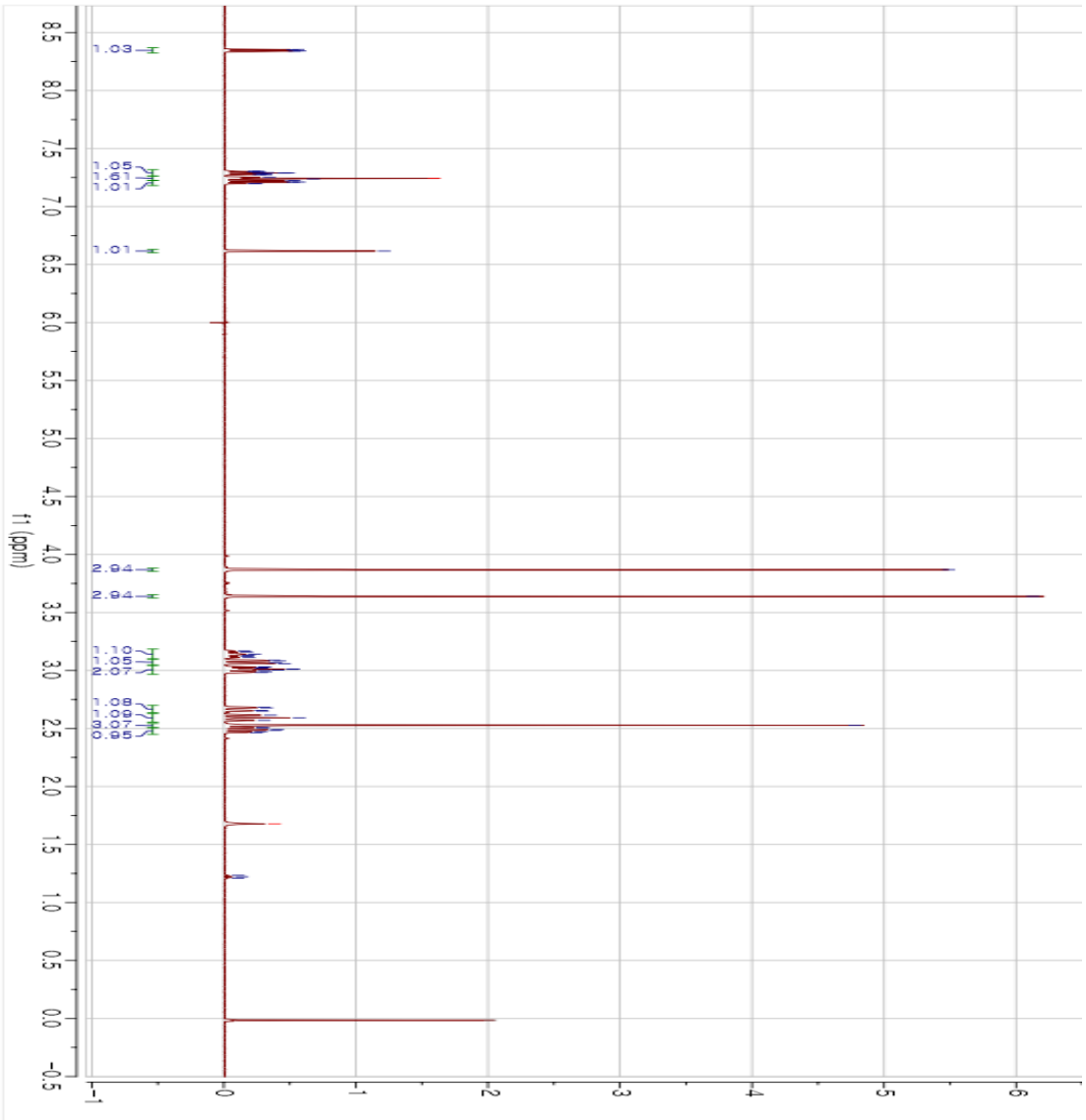


Figure 1. ^{1}H-NMR spectrum of the unknown compound. The blue numbers are the integration number of each peak.

1.6, 1.25, and 0.0 ppm (**Fig. 1**). The peak at 1.6 ppm suggests water in chloroform and at 0.0 ppm represents TMS, but the peak at 1.25 ppm are not known and ignored. The total integration number excluding the chloroform peak is around 21, suggesting the number of non-exchangeable protons.

3.1.2 Aromatic protons and protons of high chemical shifts

In the downfield, there are five protons with high chemical shifts and integration number of 1 (**Fig. 1**). The protons at 7.29, 7.24, and 7.22 ppm with similar polarity are suggested to be from an aromatic ring. Other protons that had high chemical shifts are at 8.35 and 6.62 ppm which suggest one more aromatic ring or sp^2 carbon at least. For more details, the multiplets and coupling constants of five peaks are measured with magnification and exponential function (**Fig.2**). From their multiplets and coupling constants, the structure of an di-substituted aromatic ring could be expected (**Fig. 3**). The peak of the proton at 7.22 ppm is triplet of doublet with around 7 and 2 Hz, respectively. Because the ortho-coupling and meta-coupling are typically around 7-9 Hz and around 1-3 Hz, the proton at 7.22 ppm is expected to have two protons in ortho position and one proton in meta position. Thus, the aromatic ring should have at least four protons next to each other. Furthermore, the triplet shape of ortho-coupling constant of 7-8 Hz suggests the proton is in ortho position of the proton at 7.22 ppm even though the peak of proton is a multiplet. The protons at 8.35 and 7.24 ppm have only one ortho-H and one meta-H, suggesting they are positioned in the side of the aromatic ring. The proton at 6.62 ppm is a singlet. Putting all together, the compound has one or two aromatic rings that one from 8.35-7.22 ppm is di-substituted and the other from 6.62 ppm depends on the 6 or 5 membered aromatic rings. However, 6.62 ppm can be from sp^2 carbon.

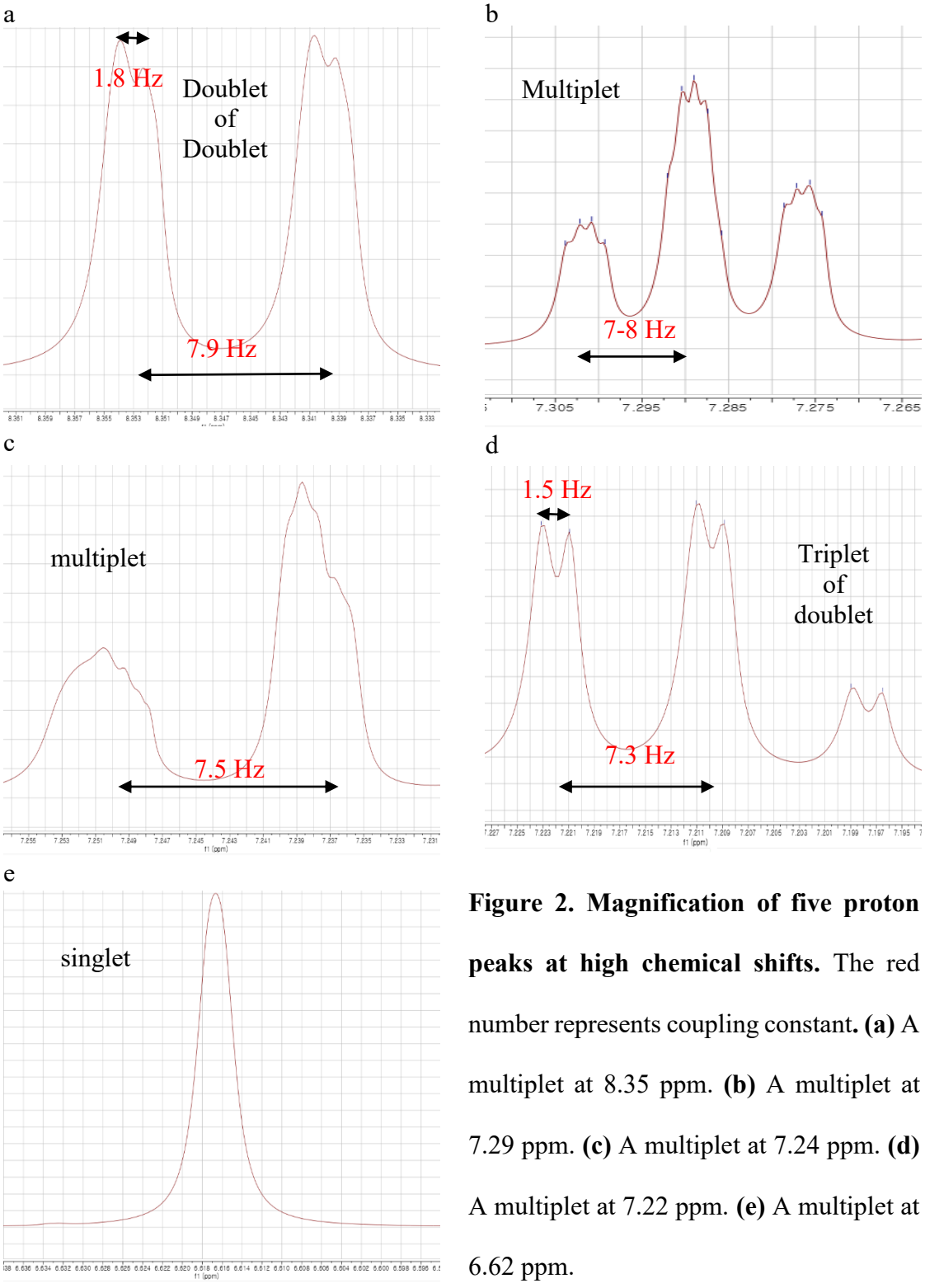


Figure 2. Magnification of five proton peaks at high chemical shifts. The red number represents coupling constant**. (a)** A multiplet at 8.35 ppm. **(b)** A multiplet at 7.29 ppm. **(c)** A multiplet at 7.24 ppm. **(d)** A multiplet at 7.22 ppm. **(e)** A multiplet at 6.62 ppm.

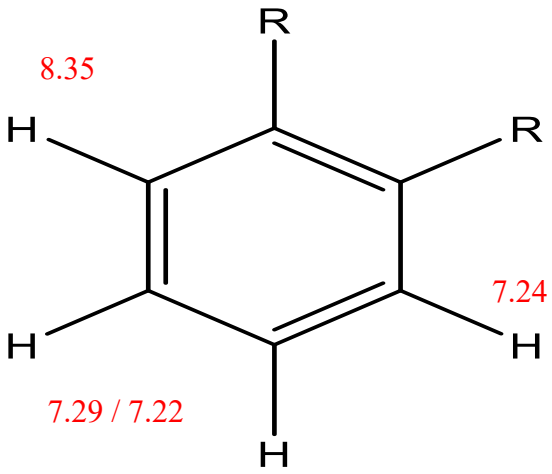


Figure 3. The expected structure of an aromatic ring. The red number represents ^{1}H chemical shift in ppm. R = any atom or group.

3.1.3 Methyl (CH3) protons

The peaks which have the integration number of around 3 are expected to represent methyl group protons. The three peaks that show the integration number are found at 3.87, 3.64, and 2.53 ppm on the chemical shift scale. Each appears as a singlet unsplit by other ^{1}H sites, (figure is not shown), and their chemical shifts are high. Thus, we anticipate that each is attached to either an O or N atom, which would explain both the downfield chemical shifts and the absence of J-splitting.

3.1.4 Methine (CH) or Methylene protons

Other protons that have integration number of 1 and 2 with chemical shifts from 3.17 to 2.49 ppm are expexted to be methine or methylene protons (**Fig. 1**). The multiplet between 3.03 and 2.98 ppm represents two protons, and the other five multiplets represent each proton. With magnification and exponential equation, the multiplicities and the coupling constants derived from each are analyzed in **Fig. 4**. The different high coupling constants of triplets of protons at 2.59 (14 Hz) and 2.49 (11.9 Hz) ppm suggest

that they are from two different methylene group: One high J-coupling comes from geminal coupling and the other comes from axial coupling in a six-membered ring.

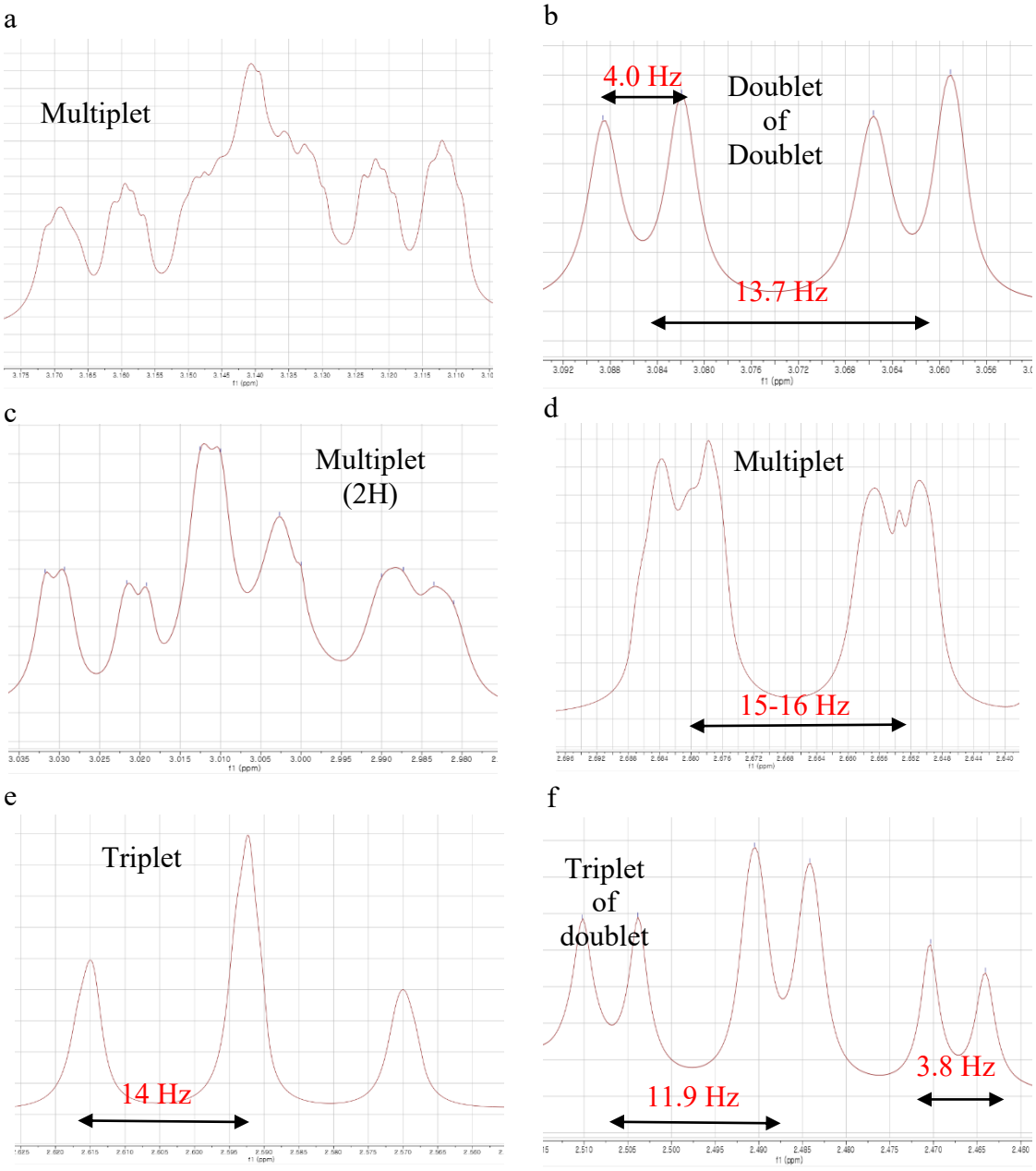


Figure 4. Magnification of six proton peaks from 3.17 to 2.49 ppm. The red number represents coupling constant**. (a)** A multiplet between 3.17 and 3.11 ppm. **(b)** A multiplet at 3.07 ppm. **(c)** A multiplet between 3.03 and 2.98 ppm. **(d)** A multiplet at 2.67 ppm. **(e)** A multiplet at 2.59 ppm. **(f)** A multiplet at 2.49 ppm

Therefore, the existence of one or two 6-membered ring is suggested.

3.2 DART-MS

Because of the complexity of assigning the connectivity of protons without the chemical formula of the unknown compound, the chemical formula of the unknown compound was required. To deduce the formula from the MS, the information from ^{1}H-NMR was used. The elution spectrum is shown in **Fig. 5a**. The highest peak at 0.97 min had a pattern of mass spectrum (**Fig. 5b**). For the formula generation, only two of the highest peaks at 296.16333 and 312.15805 m/z are used with minimum hydrogen number, 21. The atoms that were used for deducing the formulae are H, C, N, O, and P for the organic molecule and based on the isotope patterns (**Fig 6**). The formulae generated are shown

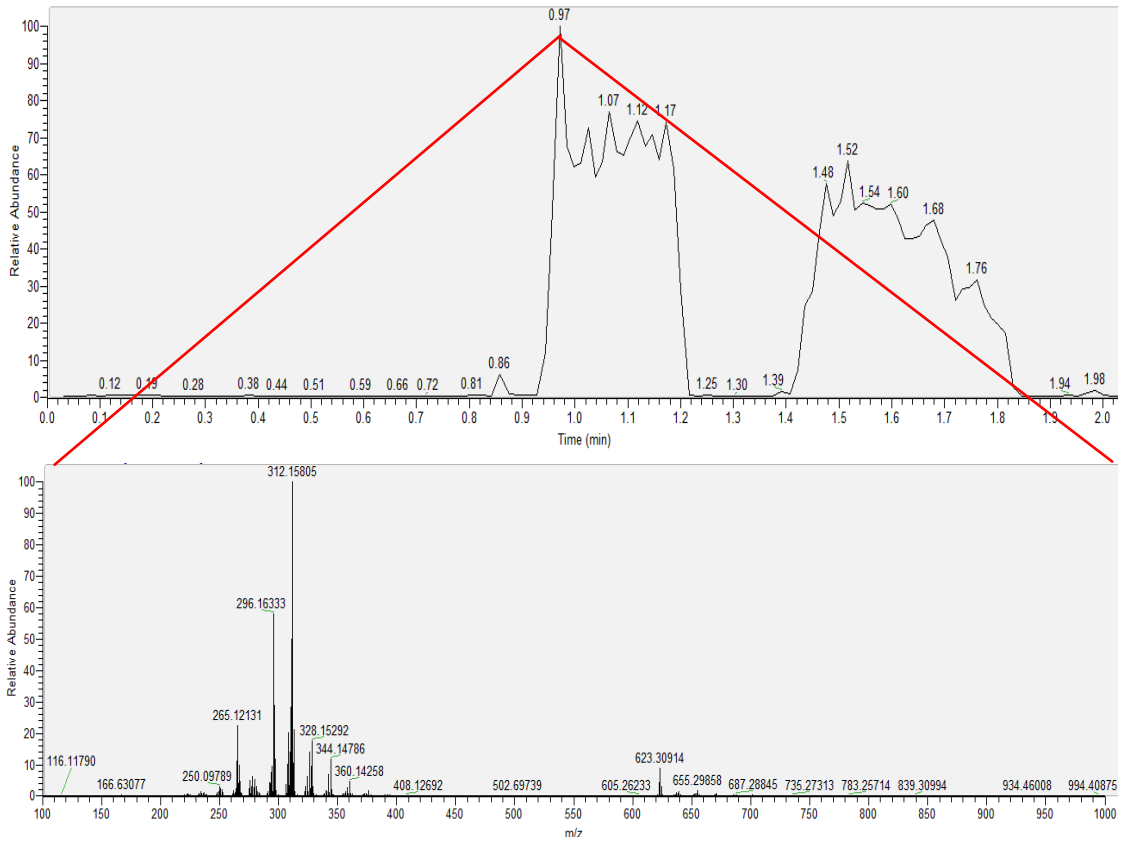


Figure 5. The elution spectrum (up) and mass spectrum at 0.97 min from the elution spectrum (down).

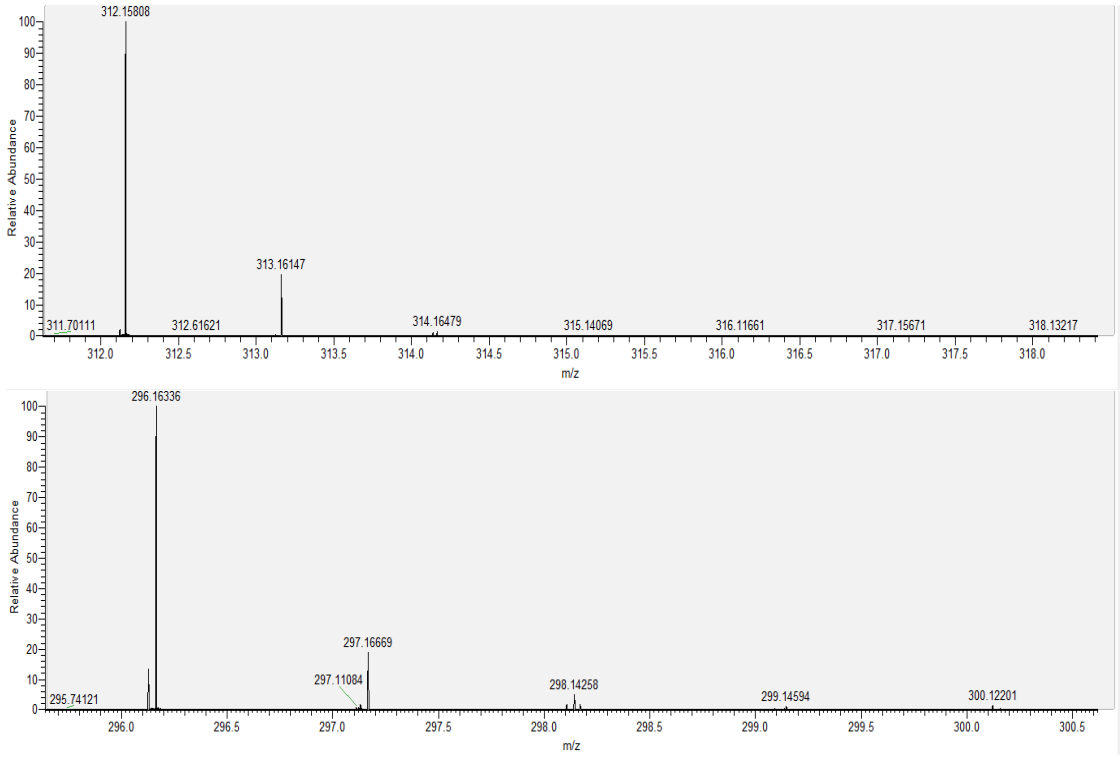


Figure 6. Isotope signature pattern of 312.15808 m/z (up) and 296.16336 m/z (down).

Idx	Formula	RDB	Delta_ppm
1	$C_{13}H_{23}ON_5P$	5.5	-0.383
2	$C_9H_{26}ON_6P_2$	1.0	-1.432
3	$C_7H_{24}N_9P_2$	1.5	3.102
4	$C_{19}H_{22}O_2N$	9.5	-3.868
5	$C_{12}H_{27}O_5NP$	0.5	4.133
6	$C_{11}H_{21}N_8P$	6.0	4.151
7	$C_{15}H_{25}O_2N_2P$	5.0	-4.916

Idx	Formula	RDB	Delta_ppm
1	$C_{13}H_{23}O_2N_5P$	5.5	-0.987
2	$C_9H_{26}O_2N_6P_2$	1.0	-1.982
3	$C_7H_{24}ON_9P_2$	1.5	2.320
4	$C_{12}H_{27}O_6NP$	0.5	3.298
5	$C_{11}H_{21}ON_8P$	6.0	3.315
6	$C_{19}H_{22}O_3N$	9.5	-4.293
7	$C_{16}H_{24}O_6$	5.0	4.293

Figure 7. The formulae generated from 296.16336 m/z (left) and 312.15808 m/z (right)

in **Fig 7**. The mass difference between 296.16333 and 312.15805 suggests that the only difference between these ions was the presence or absence of a single O atom. The RDBE should not be an integer because of the charged ion, and it should be higher than 6 at least because of the one aromatic ring, sp^2 carbon and 6-membered ring. Thus, the

result indicated that the most probable formulae of the compound are $[C_{19}H_{22}O_3N]^+$ for the 312.15805 m/z and $[C_{19}H_{22}O_2N]^+$ for the 296.16333 m/z. For the confirmation of the formulae, the isotopic signature pattern is simulated. The isotopic signature patterns that were simulated matches with the isotopic signature pattern of the sample in relative intensity. However, the parent molecule could not be determined here. The compound could be $C_{19}H_{21}O_2N$ or $C_{19}H_{21}O_3N$.

3.3 2-D NMR

3.3.1 HSQC

The connectivity between protons and carbons were determined in this section. The positively phased peaks (CH and CH_3) are shown in red color, the negatively phased peaks (CH_2) are shown in blue color, and the methyl protons show quartet of ~132 Hz of C-H coupling artifacts along the F2 (**Fig. 8**). The upfield and downfield of spectra magnified in HSQC shows the clear correlations of protons and carbons (Figures are not shown). The connectivity of protons and carbons are shown in **Table 1**. The high chemical shifts (>110 ppm) of carbons attached to the protons at 7.22-8.35 ppm confirm they are aromatic protons. In the 1D NMR section, we anticipated that the methyl groups are attached to N or O. The methyl group that has protons at 2.53 ppm is expected to be connected with N because of the lower chemical shift (44.0 ppm) of carbon than two other methyl groups (60.2 and 55.8 ppm), so the two carbons that have higher chemical shifts are attached to O. The methylene group that has protons at 2.49 and 3.02 ppm is expected to be connected with N or O because of the high chemical shift (53.3 ppm) of carbon. The methine group that has proton at 3.00 ppm is expected to be connected with N or O because of the high chemical shift (62.3 ppm) of carbon. Furthermore, the

significantly different shifts of the two methylene H atoms suggest they are close to a chiral center to make them chemically non-equivalent (diastereotopic).

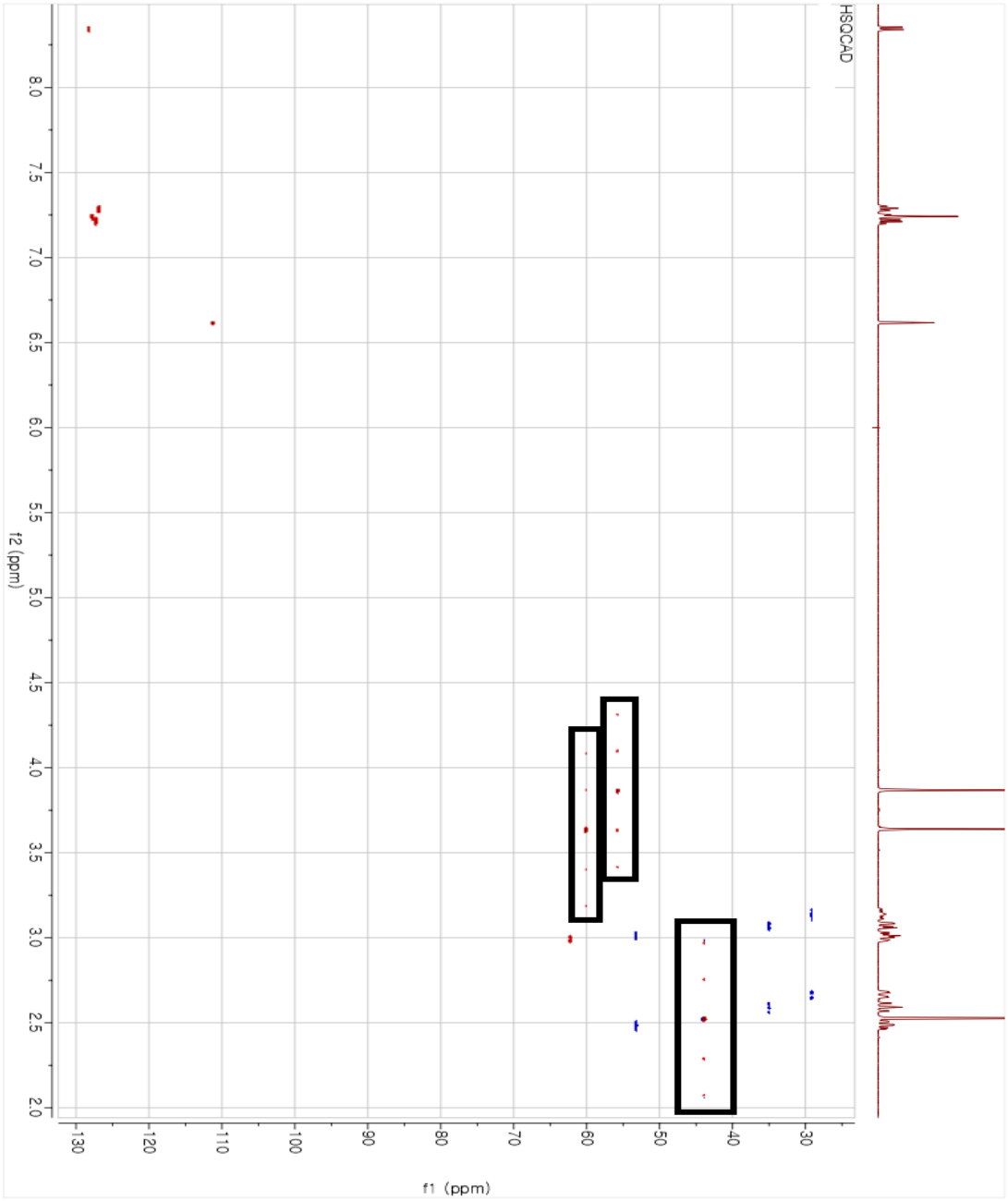


Figure 8. ^{1}H-^{13}C HSQC NMR of the unknown compound. Black rectangular represents C-H coupling artifacts along with F2 on the methyl proton peak.

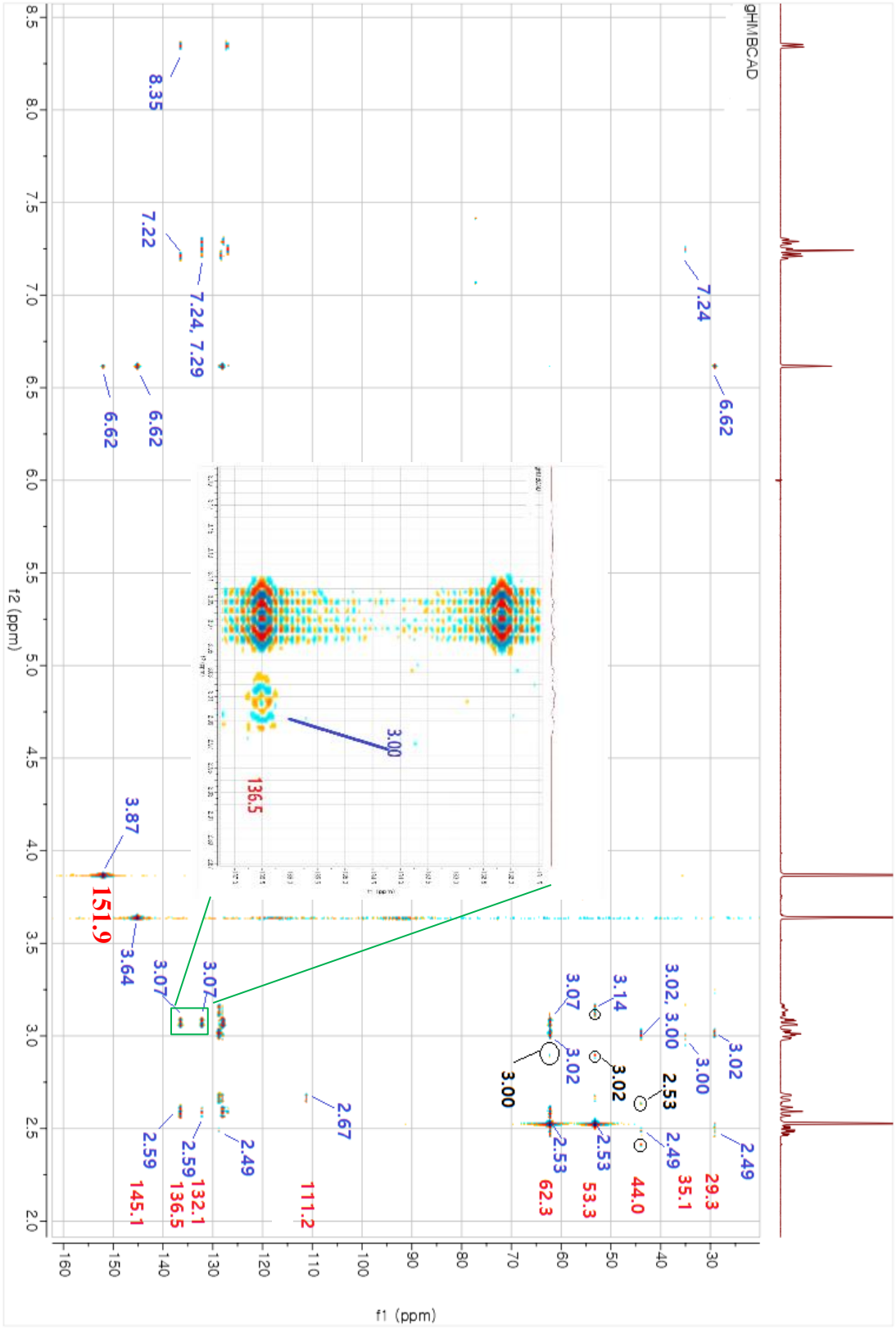
Chemical shift of H in ppm	Chemical shift of C in ppm
2.53 (3H)	44.0
3.64 (3H)	60.2
3.87 (3H)	55.8
2.49 and 3.02	53.3
2.59 and 3.07	35.1
2.67 and 3.14	29.3
3.00	62.3
6.62	111.2
7.21	127.2
7.24	127.8
7.29	127.0
8.35	128.3

Table 1. The summary of HSQC NMR result.

3.3.2 HMBC

The HMBC data showed the important information about the 2 or 3 bonds correlation between protons and carbons. Therefore, HMBC is very important for assigning the chemical structure unambiguously. The high intense peaks indicating 2 or 3 bonds were clearly shown in **figure 9a**. The peaks in F1 range between 120-130 ppm were needed to be magnified because the peaks were low-resolved (**Fig. 9b**). The correlations are summarized in **Table 2**. No carbonyl group is found, based on the highest chemical shift of carbon (<152 ppm). The two highest chemical shifts of the non-protonated carbons (151.9 and 145.1 ppm) are suggested to be double bonded carbons attached to O or N.

a


gHMBCAD
f2 (ppm)
f1 (ppm)
8.35
7.22
7.24, 7.29
6.62
6.62
6.62
7.24
3.87
151.9
3.64
3.07
3.07
3.00
136.5
3.07
3.02
3.00
3.14
3.02, 3.00
3.00
3.02
2.53
2.49
2.49
2.53
2.53
2.49
2.67
2.59
2.59
2.49
3.02
29.3
35.1
44.0
53.3
62.3
111.2
132.1
136.5
145.1

b

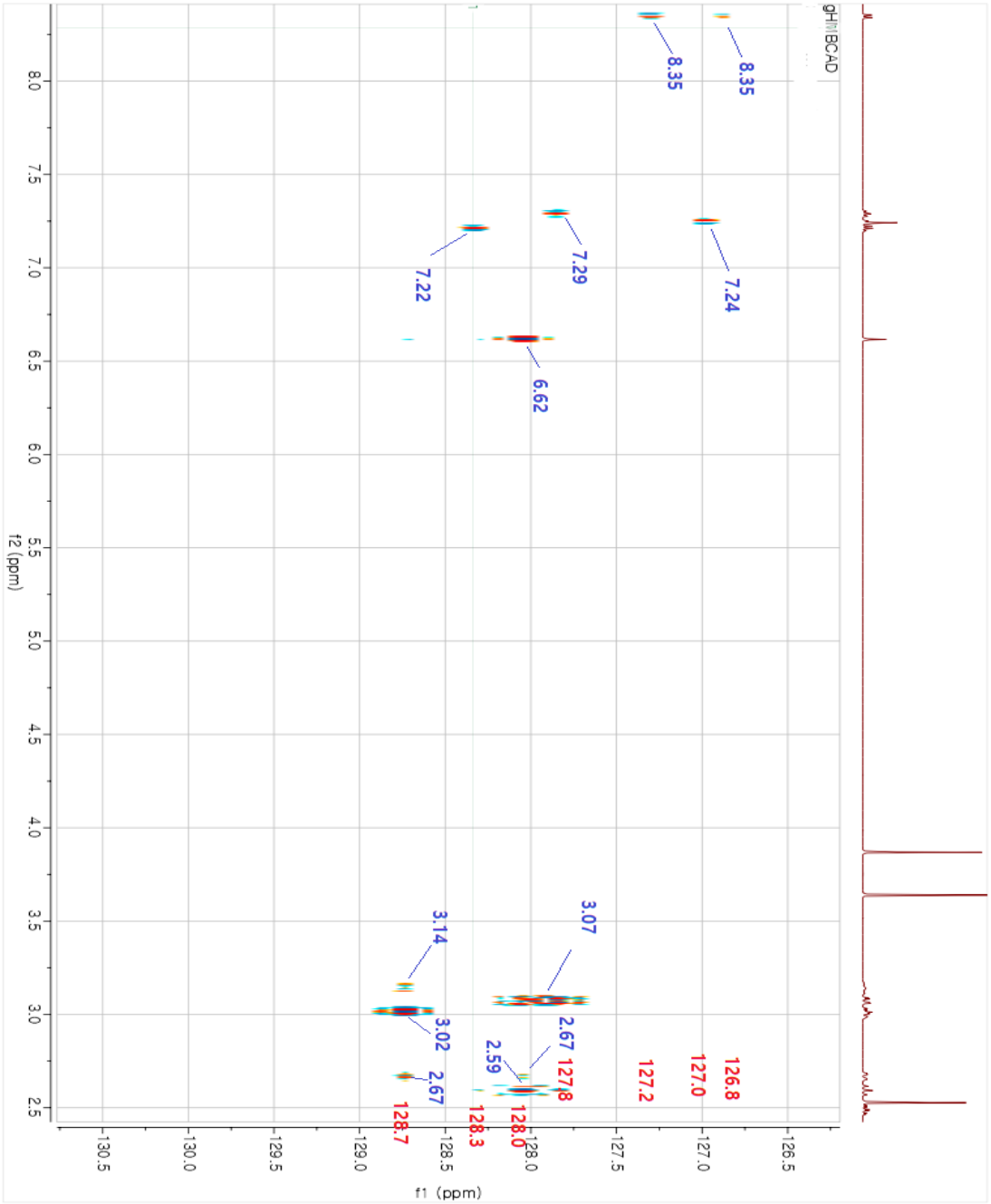


Figure 9. ^{1}H-^{13}C HMBC NMR of the unknown compound. The red number and blue number represent the ppm in F1 and F2, respectively. The black circle represents the 1 bond peaks. **(a)** The HMBC spectrum with whole range in F1 and F2. The green box region is magnified. **(b)** The HMBC spectrum with range of 130.5 to 126.5 ppm in F1 and 8.7 to 2.5 ppm in F2.

Chemical shift of H / C in ppm	Chemical shift of 2 or 3 bonded carbons
2.53 (3H) / 44.0	53.3 and 62.3
3.64 (3H) / 60.2	145.1
3.87 (3H) / 55.8	151.9
2.49 and 3.02 / 53.3	29.3, 44.0, 62.3, and 128.7
2.59 and 3.07 / 35.1	62.3, 127.8, 128.0, 132.1, and 136.5
2.67 and 3.14 / 29.3	53.3, 128.0, 128.7, and 111.2
3.00 / 62.3	35.1, 44.0, 136.5, and 128.0
6.62 / 111.2	29.3, 128.0, 145.1, and 151.9
7.22 / 127.2	128.3 and 136.5
7.24 / 127.8	35.1, 127.0, and 132.1
7.29 / 127.0	127.8 and 132.1
8.35 / 128.3	126.8, 127.2, and 136.5

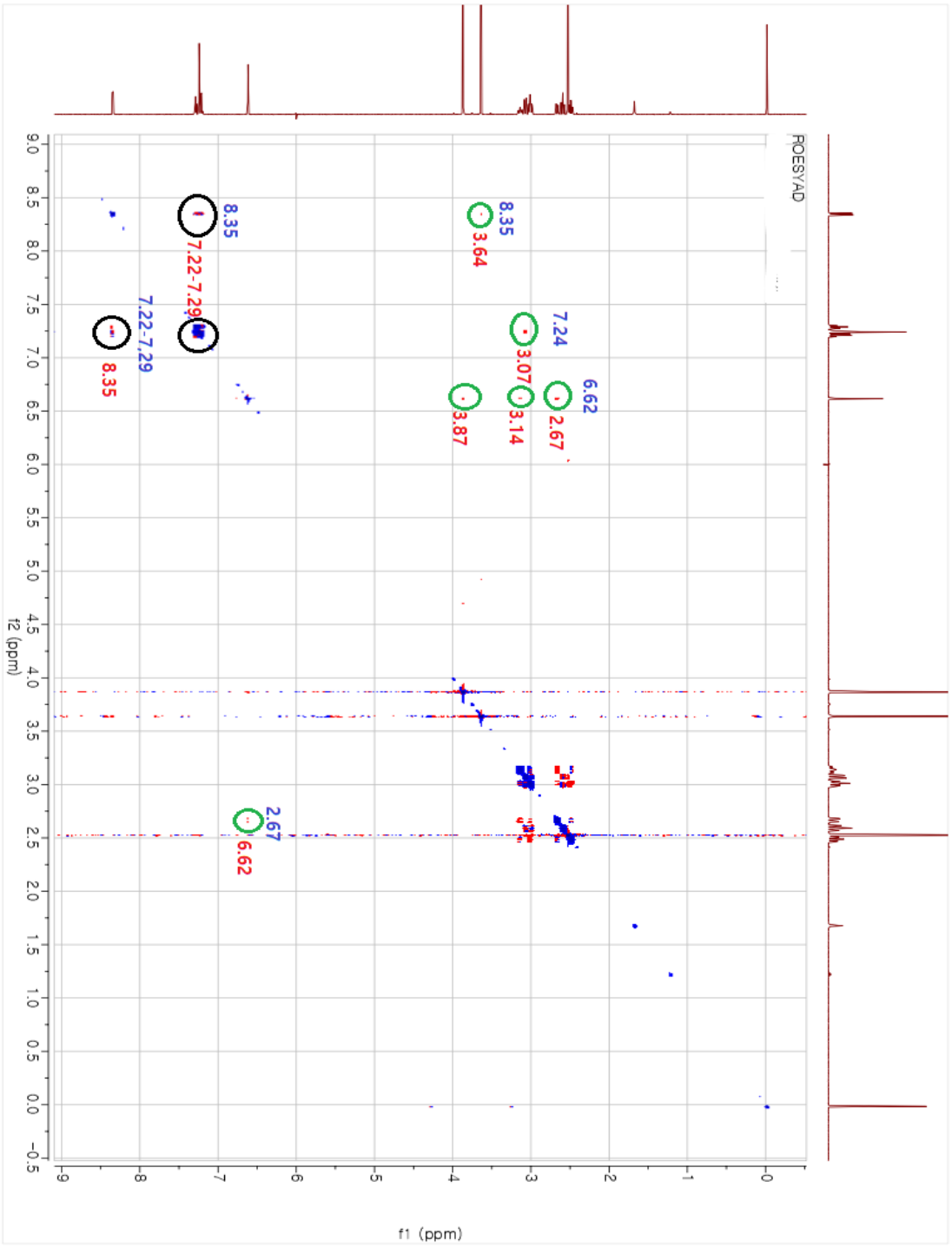
Table 2. The summary of HMBC NMR result.

3.3.3 ROESY

For assigning the stereochemistry of the unknown compound, ROESY was used, suggesting spatial proximity of protons by its cross peaks. There are two different cross peaks in the spectrum (**Fig. 10**). The ROE should be phased in opposite to the diagonal peak whose color is in blue. Thus, the red color represents ROE and the mixed color in red and blue were suggested to be artifacts, possibly TOCSY / COSY-like[7] (**Fig. 10a**). However, the type of artifacts was unsure because we failed to see COSY results. Thus, the cross peaks that have mixed color are ignored. In the range 2.49-3.16, the peaks were so messy, so the range was magnified (**Fig. 10b**). The peaks in the range of 2.49-2.59 that are not clearly resolved were not determined. The proximity of protons is

summarized in **Table 3**. There was no chemical exchange, which normally pure blue cross-peak represents, observed.

a



b

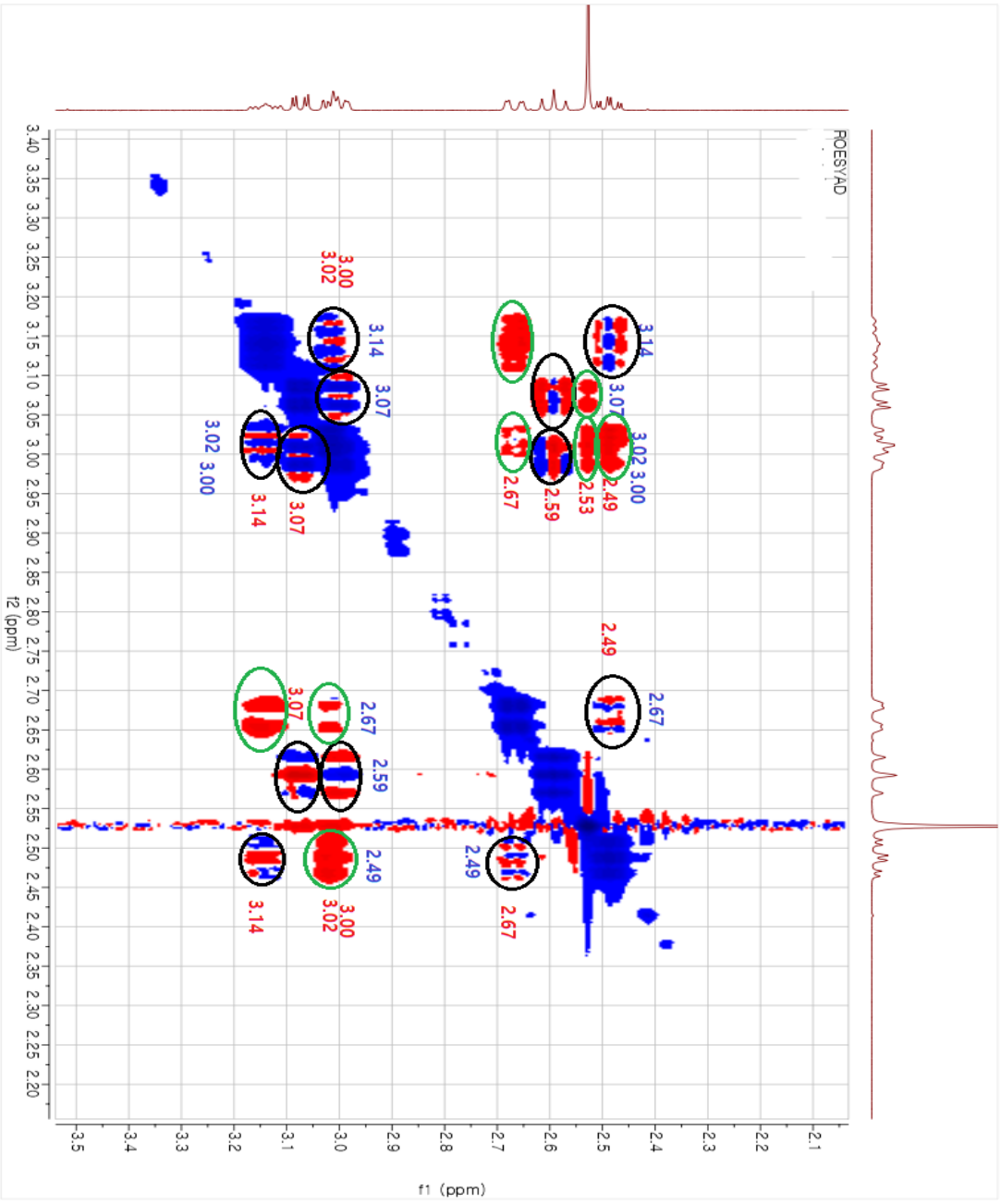


Figure 10. The ^{1}H-^{1}H ROESY NMR spectrum. The green circle represents ROE corss peak. The black circle represents TOCSY-like artifacts. **(a)** The ROESY spectrum with whole range. **(b)** The ROESY spectrum with range of 3.5 to 2.1 ppm in F1 and 3.4 to 2.2 ppm in F2.

Chemical shift of H in ppm	Chemical shift of proximal H in ppm
2.53 (3H)	3.00, 3.02, and 3.07
3.64 (3H)	8.35
3.87 (3H)	6.62
2.49	3.00
3.02	2.53, and 2.67
2.59	-
3.07	2.53 and 7.24
2.67	3.02 and 6.62
3.14	6.62
3.00	2.53 and 2.49
6.62	2.67, 3.14, and 3.87
7.22	-
7.24	3.07
7.29	-
8.35	3.64

Table 3. The summary of ROESY NMR result.

3.4 COMPREHENSIVE RESULTS

Before starting structure elucidation with HMBC and ROESY, what has been known from each result is summarized for helping the elucidation. **Table 4** shows the summary of all NMR results.

The chemical formula of the unknown compound is $C_{19}H_{21}NO_3$ or $C_{19}H_{21}NO_2$. Because of the number of all the non-exchangeable protons is 21 which is exactly matched with the number of protons in the formula, the formula suggests that there is

no -OH. Furthermore, we anticipated that there is no carbonyl. Thus, all O is expected to be from ether group.

One aromatic ring and one sp^2 carbon or another aromatic ring are suggested: one di-substituted aromatic ring contains protons at 8.35, 7.29, 7.24, and 7.22 ppm, and the other aromatic ring or sp^2 carbon contains proton at 6.62 ppm. The two methyl groups (C-14 and C-15) that contain protons at 3.87 and 3.64 ppm are suggested to be attached to O, and the other methyl group (C-17) that contain proton at 2.53 ppm are suggested to be attached to N. The two high coupling constants of two methylene group (C-16 and -18) protons at 2.59 and 2.49 ppm are suggested to be in axial position on 6-membered ring. Furthermore, the methylene group (C-16) that contains proton at 2.49 ppm is suggested to be attached to N or O. The methine group (C-13) that contains proton at 3.00 ppm is suggested to be attached to N or O.

All the information is suggestive and ambiguous. In the next step, using HMBC and ROESY results, the structure will be determined.

#	δ C (ppm)	δ H (ppm)	H-H Coupling constants (Hz)	HMBC (#)	ROESY (#)
1	151.9	-	-	-	-
2	145.1	-	-	-	-
3	136.5	-	-	-	-
4	132.1	-	-	-	-
5	128.7	-	-	-	-
6	128.3	8.35	7.9 and 1.8	3, 9, 11	14
7	128.0	-	-	-	-
8	127.8	7.24	7-8	4, 10, 18	18b

9	127.2	7.22	7.3, 7.3, and 1.5	3, 6	-
10	127.0	7.29	7-8	4, 8	-
11	126.8	-	-	-	-
12	111.2	6.62	-	1, 2, 7, 19	15, 19a, 19b
13	62.3	3.00	-	3, 7, 17, 18	16a, 17
14	60.2	3.64 [3H]	-	2	6
15	55.8	3.87 [3H]	-	1	12
16-a	53.3	2.49	11.9, 11.9, and 3.8	5, 13, 17, 19	13
16-b		3.02	-		17, 19a
17	44.0	2.53 [3H]	-	13, 16	13, 16b, 18b
18-a	35.1	2.59	14 and 14	3, 4, 7, 8, 13	-
18-b		3.07	13.7 and 4.0		8, 17
19-a	29.3	2.67	15-16	5, 7, 12, 16	12, 16b
19-b		3.14	-		12

Table 4. The summary of comprehensive NMR results.

CHAPTER 4

STRUCTRE ELUCIDATION

With all information from the HMBC and ROESY, the structure elucidation was obtained step by step. The easiest way for the elucidation is to start with terminal group of a molecule, which is the methyl group (**Fig 11a**). Beginning with the H-17 methyl protons, the protons are three bonds away from the C-13 and -16 which is connected to H-13 and -16 through N. The C-16 is suggested to be 6-membered ring. Starting again with the H-16, H-16 is correlated with C-5 and -19. Because of the 6-membered ring, the C-5 which is considered to be double bonded has to have three bonds correlation with H-16. Therefore, the C-19 has two bonds correlation. From H-19, it is correlated with C -7 and -12. The C-12 is considered to be sp^2 carbon and protonated as we analyzed. It should be put on other places rather than the 6-membered ring because H-12 is correlated with C-1 and -2. If C-12 is located on the 6-membered ring, there will be impossible C-1 and -2 correlations. The positions of C-1 and -2 cannot be determined.

The protons which is correlated with C-1 and -2 are H-14 and -15 which are attached to the remaining two methyl groups. Beginning with H-15 and -14, the three bonds correlation with C-1 and -2 gives the structure shown in **Fig. 11b**. The merged structure is shown in **Fig. 11c**. The aromatic ring that was expected from previous analysis is also terminal part of the structure (**Fig. 3**). H-8 and H-13 share the C-18 correlation, H-9 and H-13 share the C-3 correlation, and H-10 has C-4 correlation. Furthermore, because C-18 is on 6-membered ring as it was analyzed, there are only three possible structures. However, when C-3 is located next to the C-7, the H-6 correlation with C-11 is impossible (**Fig. 12a**). Moreover, when C-3 is two bonds away

from H-13, the H-13 and C-18 correlation, and H-6 and C-11 correlation will be impossible (**Fig. 12b**). Thus, only one possible structure remains and meets all the requirement (**Fig. 12c**).

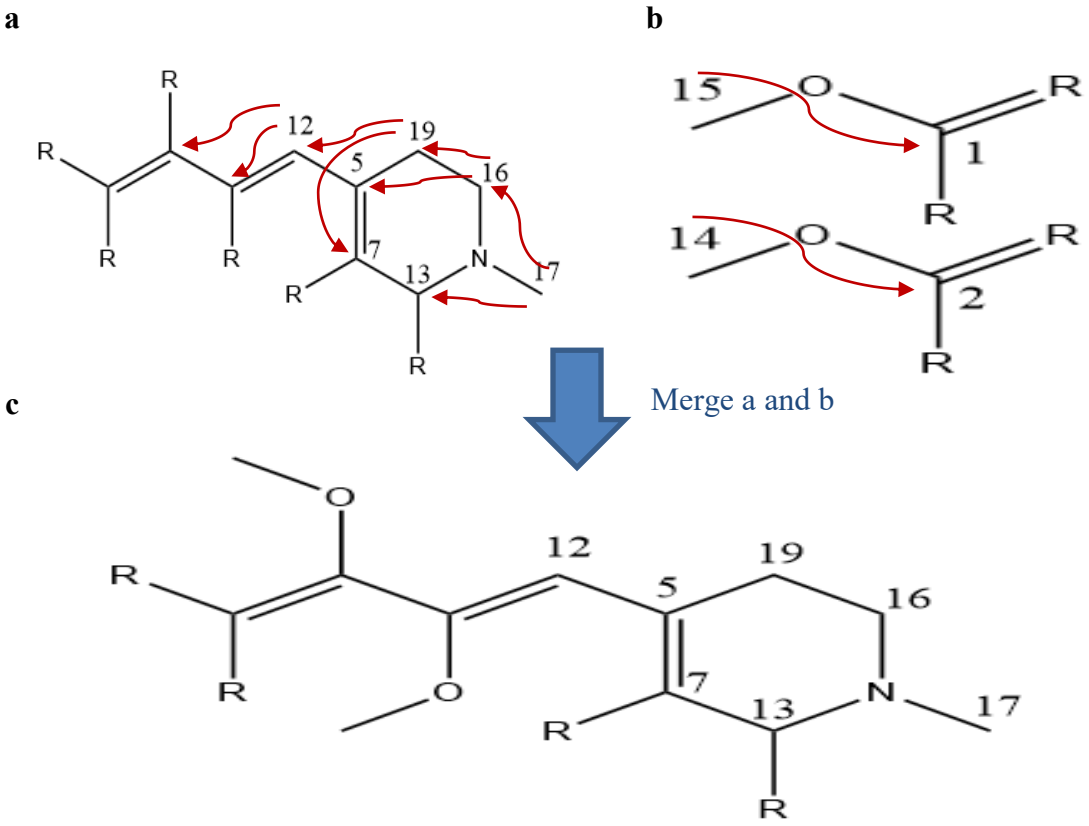


Figure 11. Partial structure step 1. Red arrow represents HMBC correlation. The letter R represents arbitrary atoms or groups. **(a)** The structure from H-17. **(b)** The structure from H-14 and -15. **(c)** The structure that is merged from **a** and **b**.

The carbons are all fulfilled in the partial structure. The only possible way to match the number of carbons (19) is to merge the two branches of non-identified sp^2 carbons. To specify the location of C-1 and -2, the ROESY results (H-15 and H-12 cross

peak, and H-14 and H-6 cross peak) were used. The proposed structure, which meets all the requirement, is shown in **Fig. 13**.

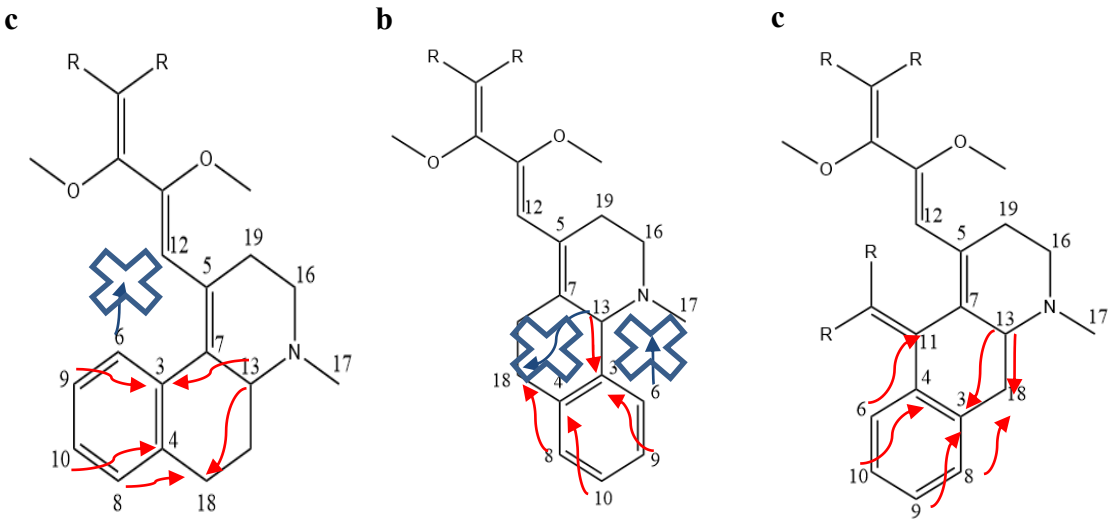


Figure 12. Partial structure step 2. Red arrow represents HMBC correlation. The letter R represents arbitrary atoms or groups. **(a, b)** The incorrect structure of the molecule. **(c)** The partial structure.

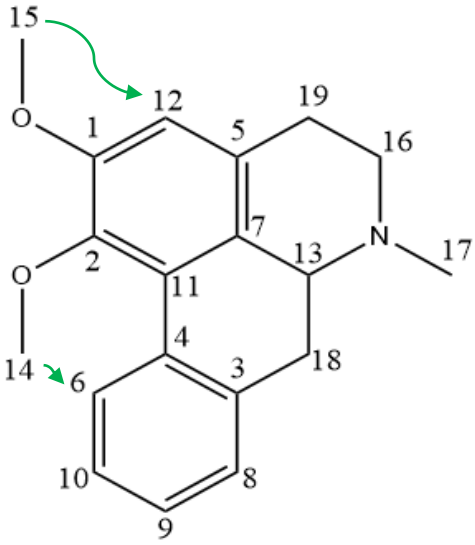
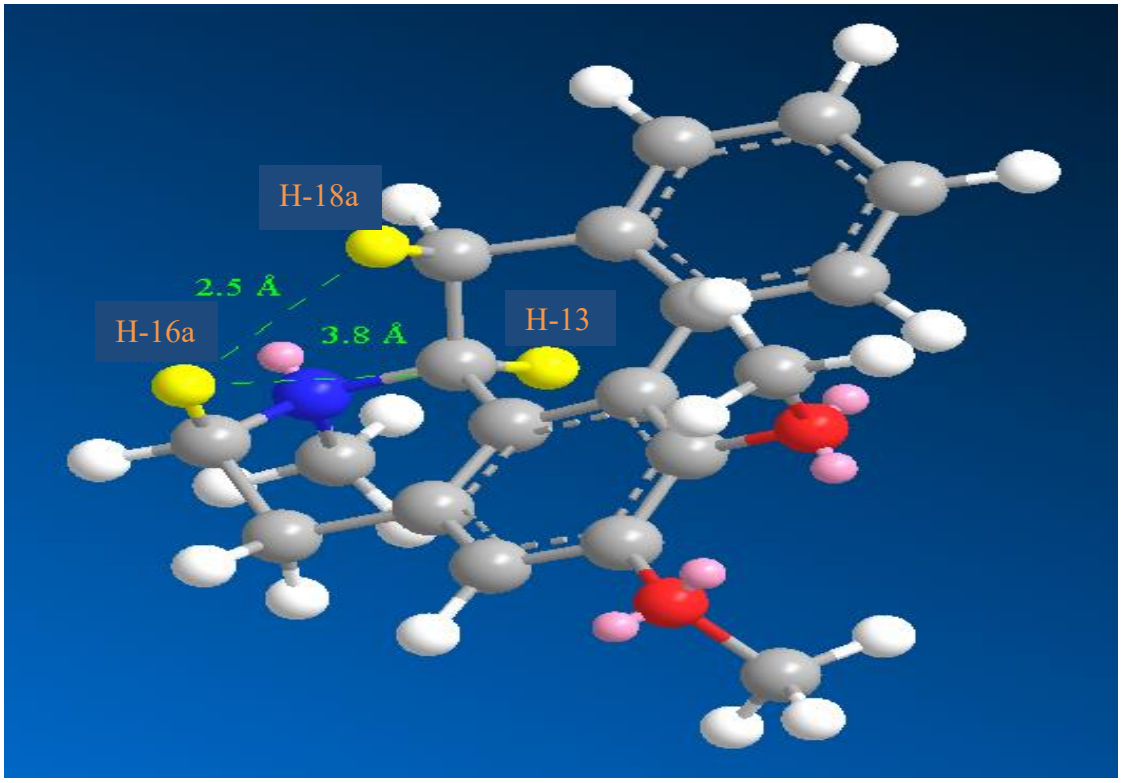


Figure 13. The proposed structure. Green arrow represents the ROE.

Exact position of methylene protons could be determined by their coupling constants as mentioned in the previous section. With their assignment, the ROE between H-13 and H-16, and the absence of ROE between H-16 and H-18 give clues as to the

stereochemistry. If the chiral center (C-13) were in S configuration, the ROE between H-16a (2.49 ppm) and H-18a (2.59 ppm) should have been observed because of the lower distance (2.5 Å) than the distance (3.8 Å) between H-13 (3.00 ppm) and H-16a (2.49 ppm) (**Fig. 14a**). Thus, the chiral center is R configuration, and the ROE from H-17 also suggests C-17 is in equatorial position (**Fig. 14b**). Putting all together, the chemical structure of the unknown molecule was deduced, and all protons and carbons are assigned (**Fig. 15 and Table 5**). The unknown molecule is determined as (6aR)-1,2-dimethoxy-6-methyl-5,6,6a,7-tetrahydro-4H-dibenzo[de,g]quinoline or nuciferine.

a



b

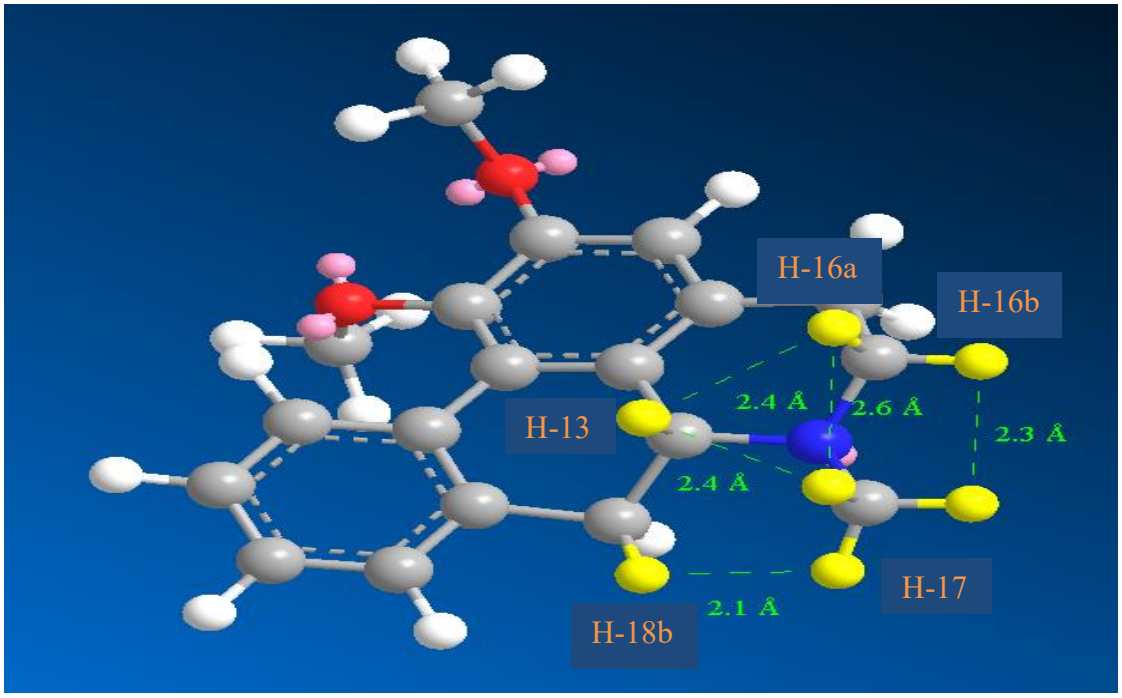


Figure 14. The 3D structure of the compound. The 3-D structure is created, and the distance between protons is measured by Chem 3D. The orange letter represents the proton number. The green letter and dash represent the distance between protons. Yellow atom represents the proton corresponded to the proton number **(a)** The 3-D structure of S configuration. **(b)** The 3-D structure of R configuration.

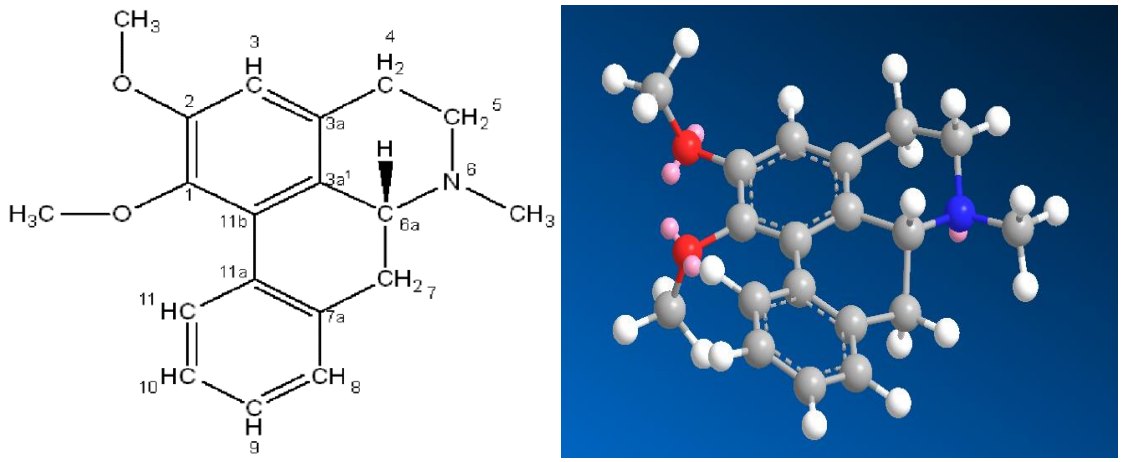


Figure 15. The proposed final structure. The 3-D structure is created by Chem 3D. The proposed final 2-D structure (left). The proposed final 3-D structure (right).

Carbon Assignment					
C-1	145.1	C-6a	62.3	C-11a	132.0
C-2	151.9	C-7	35.1	C-11b	126.8
C-3	111.2	C-7a	136.5	NCH_3	44.0
C-3a	128.7	C-8	127.8	1-OCH_3	60.2
C-3a1	128.0	C-9	127.2	2-OCH_3	55.8
C-4	29.3	C-10	127.0		
C-5	53.3	C-11	128.3		
Proton Assignment					
C-3-H	6.62	C-8-H	7.24	1-O-C-3H	3.64
C-4-2H	2.67, 3.14	C-9-H	7.22	2-O-C-3H	3.87
C-5-2H	2.49, 3.02	C-10-H	7.29		
C-6a-H	3.00	C-11-H	8.35		
C-7-2H	2.59, 3.07	N-C-3H	2.53		

Table 5. The assignment of chemical shifts of carbons and protons.

CHAPTER 5

DISCUSSION AND CONCLUSION

The chemical structure of the unknown compound was determined with MS-NMR combined technique. The results that were consistent with the structure confirm the unambiguousness of the structure, except for the stereochemistry.

The compound (295 m/z) is relatively small sized, compared to other large molecules (>1000 m/z), and because of that, the NMR spectra were not very complex. Furthermore, the NMR spectra were well resolved due to the high resolution (600 MHz). Because of the simplicity and high resolution, the structure could be determined without COSY spectrum. However, because of the absence of COSY, it took more time to find additional information about the structure. For example, if COSY data was ready to be used, the assignment of methylene and methine protons would be much faster. In this thesis, the necessity of COSY was realized.

The biggest question from this thesis came from the MS data. After determining the structure, $C_{19}H_{21}NO_2$ (295 m/z) was determined for the chemical formula of the compound. However, the highest abundance of ion eluted from MS was $[C_{19}H_{22}NO_3]^+$ (312 m/z). This means the ionic form of the compound was dominated in $[M+OH]^+$, and this result is still nonsense to me. How the environment of DART-orbitrap creates OH adduct form from the compound, and what is the structure of the adduct form are challenging to me. Because of the result, the formula of the compound was thought to be $C_{19}H_{21}NO_3$ at the first, and it made me waste time for a long time because of the inconsistence with the NMR data. From this, the fact that the MS alone can hardly determine a chemical formula of an unknown compound is learned.

The nuciferine is almost plane-like structure because of the two aromatic rings attached to one 6-membered ring. Because of this property, determining the stereochemistry was very difficult. Even though the configuration of the chiral center in the molecule changes, the structure does not change much. Thus, the stereochemistry of the structure was determined by one miss cross-peak, rather than enough information from ROESY. For more certainty, ROESY or NOESY experiment should be iterated.

REFERENCES

1. J.H. Gross. Direct analysis in real time – a critical review on DART-MS *Anal. Bioanal. Chem* 2014, 406, 63-80.
2. A.D. Lesiak, J.R.E. Shepard. Recent advances in forensic drug analysis by DART-MS *Bioanalysis* 2014, 6, 819-842.
3. Q. Hu, R.J. Noll, H. Li, A. Makarov, M. Hardman, R.G. Cooks. The orbitrap: a new mass spectrometer *J. Mass Spectrom* 2005, 40, 430-443
4. R.A. Zubarev, A. Makarov Orbitrap mass spectrometry *Anal. Chem* 2013, 85, 5288-5296.
5. T.-L. Hwang and A.J. Shaka. Reliable two-dimensional rotating-frame cross-relaxation measurements in coupled spin systems *Journal of Magnetic Resonance Series B* 1993, 102, 155-165.
6. M.J. O'Neil, P.E. Heckelman, C.B. Koch, K.J. Roman, The Merck Index, an Encyclopedia of Chemicals, Drugs, and Biologicals – Fourteenth Edition, Merck Co., Inc. Whitehouse Station, NJ 2006.
7. Neuhaus D., Keeler J. "False" Transverse NOE Enhancements in CAMELSPIN Spectra *J. Magn. Reson* 1986, 68, 568.